* The first assumption here is that the csv files are in the right directory.
* In terms of the code, the “programA”, “programB”, and “chosen\_area” don’t have to be parametric. The code could easily divide the data into programs and areas, without getting any input, and analyse the biometric data itself.
* By looking at all the districts simultaneously, the overlap metric would be more explanatory. This is because currently, we can only see the overlap metric per district, not the whole program. By analysing all of the data, we would have a better understanding of how much the programs overlap.
* After assessing the overall performance of the overlaps in two companies, the code can also review different overlaps for different areas by itself, without getting any inputs from the user. It is just a matter of getting the unique area names from the “Areas” column of the “beneficiaries.csv”.
* The code can be made more efficient by combining several lines. For example, “all\_ids\_in\_A” and “chosen\_ids\_in\_A” can be combined in a single line with an “&” operator in the filter function.